

Therefore, although this invention has been described with a certain degree of particularity, it is to be understood that the present disclosure has been made only by way of illustration, and that numerous changes in the details of construction and arrangement of components may be resorted to without departing from the spirit and scope of the invention.

REFERENCES

- Baayen, H., van Halteran, H., Neijt, A., Tweedie, F. (2002). "An Experiment in Authorship Attribution." *Journées internationales d'Analyse statistique des Données Textuelles* 6.
- Chaski, C. E. (1997). "Who Wrote It? Steps Toward A Science of Authorship Identification." *National Institute of Justice Journal*: 15-22.
- Chaski, C. E. (2001). "Empirical Evaluations of Language-Based Author Identification Techniques." *Forensic Linguistics* 8(1): 1-65.
- Chaski, C. E. (2004). "Recent Validation Results for the Syntactic Analysis Method for Author Identification." International Conference on Language and Law, Cardiff, Wales.
- Hilton, O. (1993). *Scientific Examination of Questioned Documents*. Boca Raton, Fla., CRC Press.
- Koppel, M, Schler, J. (2003). "Exploiting Stylistic Idiosyncrasies for Authorship Attribution", in *Proceedings of IJCA'03 Workshop on Computational Approaches to Style Analysis*", Acapulco, Mexico.
- Manning, C. D. and Schuetze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Mass.: MIT Press.
- McMenamin, G. R. (2003). *Forensic Linguistics; Advances in Forensic Stylistics*. Boca Raton, Fla., CRC Press.
- Miron, M. S. (1983). "Content Identification of Communication Origin." *Advances in Forensic Psychology and Psychiatry*. R. W. Reiber. Ed. Norwood, N.J., Ablex.
- Mosteller, F., Wallace, D. L. (1984). *Applied Bayesian and Classical Inference: The Case of the Federalist Papers*. New York, Springer-Verlag.
- Stamatatos, E., Fakotakis, N., Kokkinakis, G. (2000). "Automatic Text Categorization in Terms of Genre and Author." *Computational Linguistics* 26(4): 471-495.
- Stamatatos, E., Fakotakis, N., Kokkinakis, G. (2001). "Computer-Based Authorship Attribution Without Lexical Measures." *Computers and the Humanities* 35: 193-214.
- Tambouratzis, G., Markantonatou, S., Hairetakis, N., Vassiliou, M., Carayannis, G., Tambouratzis, D. (2004). "Discriminating the Registers and Styles in the Modern Greek Language—Part 2: Extending the feature Vector to Optimize Author Discrimination." *Literary & Linguistic Computing* 19(2): 221-242.

I claim:

1. A computer-aided method to determine whether an unidentified author of a textual work belongs to a group comprising the textual work of a known author, the method comprising the steps of:

- obtaining a sample of the textual work of the unidentified author;
- obtaining a sample of the textual work of the known author;
- entering the samples into a computer system, the computer system including a memory, a means for analyzing documents, and a means for determining belonging, stored within the memory;

utilizing the means for analyzing documents, splitting the entered samples into individual sentences, the sentences each including a head, a plurality of words and punctuation, the punctuation defining a syntactic edge within the individual sentence, and graphemic features defining a discursive function emphatic selected from the group consisting of: bolding, italics, capitalization, emoticons and serial punctuation marks;

categorizing the punctuation by determining the syntactic edge;

indicating the discursive function emphatic, a graphemic feature being generated by the steps of categorizing and indicating;

dividing each of the individual sentences into the words;

labeling each of the words as a part of speech;

listing the labeled words into phrases for each labeled word;

identifying phrases for each said head;

classifying the identified phrases as marked or unmarked;

characterizing the identified phrases by markedness, thereby producing a plurality of syntactic features; and

utilizing the means for determining belonging, inputting at least one of the syntactic features and inputting at least one feature selected from the group consisting of: (i) the graphemic features and (ii) the syntactic edges defined by punctuation; wherein the inputting is done for each said sample to determine whether the unidentified author of the textual work sample belongs to the known author group.

2. A system for determining whether an unidentified author of a textual work belongs to a group comprising the textual work of a known author, the system comprising:

- a computer system including a memory, an input means, a means for analyzing documents, and a means for determining belonging, stored within the memory;
- a sample of the textual work of the unidentified author;
- a sample of the textual work of the known author, the samples being input into the computer system;

the means for analyzing documents splitting the entered samples into individual sentences, the sentences each including a head, a plurality of words and punctuation, the punctuation defining a syntactic edge within the individual sentences, and graphemic features defining a discursive function emphatic selected from the group consisting of: bolding, italics, capitalization, emoticons and serial punctuation marks; the means for analyzing documents categorizing the punctuation by determining the syntactic edge; and indicating the discursive function emphatic, thereby generating a graphemic feature;

the means for analyzing documents dividing each of the individual sentences into the words; labeling each of the words as a part of speech; listing the labeled words into phrases for each labeled word, identifying phrases for each said head, classifying the identified phrases as marked or unmarked, characterizing the identified phrases by markedness, thereby producing a plurality of syntactic features; and

inputting at least one of the syntactic features and inputting at least one feature selected from the group consisting of (i) the graphemic features and (ii) syntactic edges defined by punctuation; wherein the inputting is input into the means for determining